www.icgst.com

Rotational invariant Real Time Text Recognition

M. Anyayahan, M. Balinas, A. La Madrid, M. Laurel, C. Lopez, R. Tolentino
Electronics Engineering, Polytechnic University of the Philippines – Santa Rosa Campus, Laguna, Philippines
marron.jann.21@gmail.com balinaskelaine@gmail.com aldrinlamadrid@gmail.com
mark.laurel1595@gmail.com teljoylopez@gmail.com kenmetara@yahoo.com
http://www.pup.edu.ph

Abstract
In everyday life, people always encounter different text images. These text images are in a style of linear or multi-oriented texts in either printed or written form. Due to different orientations of texts in an image, it is a challenge in Optical Character Recognition to recognize this kind of text. In this paper, real time recognition of text in different rotational variations is presented. The performance is done from acquisition of image by a camera and processed by Microsoft Visual Studio. The detection and recognition of text with different rotational variations are achieved by detecting and computing the direction and angle of tilt respectively through the use of geometric and trigonometric principles then recognized by Tesseract optical character recognition engine after counter rotation.[1]

Keywords: *multi orientation angle, rotational variation, tilt angle, tilt direction, Optical Character Recognition.*

Nomenclature

OCR	Optical Character Recognition
BLOB	Binary Large Object
ROI	Region of Interest
CC	Character Confidence
WC	Word Confidence

1. Introduction
Text is a human-readable sequence of characters and the words they form are in either written or printed work. These characters are often in the form of alphanumeric that created series of words. Reading text is a part of our everyday lives. But these texts are not always in a horizontal manner that humans usually see and easily read. Different orientations of text existed due to the creativity of humans, and these text arranged in different orientations can also be certainly read by humans because of their perception. But detection and recognition of these texts in different orientations is a challenge in the field of machine vision.

Numerous studies have been conducted to advance the recognition of text in multi-orientation. The study focuses on end-to-end real-time text localization and recognition method. They present that the real-time performance is achieved by posing the character detection problem as an efficient sequential selection from the set of Extremal Regions. All of the features are scale-invariant, but not all are rotation-invariant however, the features are somewhat robust against small rotations [3]. Another is a proposed technique to extract text from natural scene images but the proposed system is sometimes not able to detect and extract text properly because of some factors like the image may be tilted, some shadow area or the background is complex [1].

A recent study entitled Text-Line Detection, Segmentation and Recognition in Natural Scene presented scene text detection and extraction from images and an algorithm which involves pre-processing of images by applying wiener filter and run length method to detect the text in images. This algorithm does not only detect the text in image but it also detects the blur text. The problem with this study is the certain limitations stated that text with multi-orientation angle cannot be detected [4].

To solve this problem, a system is proposed by the researchers to recognize text with different rotational variations by detecting and computing the direction and angle of tilt respectively through the use of some geometric and trigonometric principles then implementing Optical Character Recognition after counter rotation.

This research is essential to aid the existing studies in advancing image processing. The vital part is to make it more efficient to read text on different rotation variations and to present a new method for detecting tilt direction and angle in text characters. This can also be useful in further studies or development of study about tilt direction and tilt angle in character recognition.

The remainder of the paper is organized as follows: Section (2) focuses on how the system is implemented and evaluated. Section (3) emphasize the results in identifying the direction and angle of tilt and the evaluation result of the system's reliability.

2. Theory
The system is implemented and evaluated using a computer with an Intel core i3 (2GHz) microprocessor with 4 GB RAM running at Windows 10 Home 64-bit Edition, and the sensor used is A1 Tech AW-06 Webcam

with a resolution of 640x480 pixels (30 frames per second). The starting distance is 35 cm. The samples are printed in a 8.5"x11" in Calibri font style and 72 font size. Figure 1 shows the whole process of the system. Text image acquired by the camera is considered as the input image and the one that will be processed by the system.

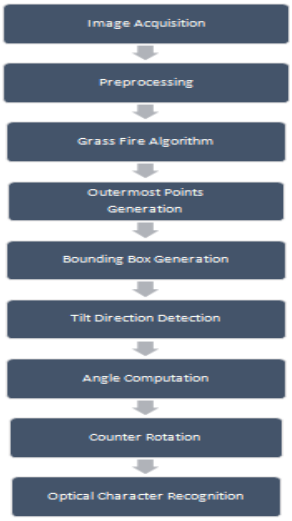


Figure 1 Conceptual Framework

Figure 1 shows the whole process of the system. Text image acquired by the camera is considered as the input image and the one that will be processed by the system. After the image has been acquired, it will be subjected to image pre-processing. Pre-processing includes image binarization and canny edge detection. Image pre-processing is done to make the image free to noise and be converted to full binary image. After pre-processing, Grass Fire Algorithm will be implemented. Grass fire Algorithm works by burning the pixel from a certain point or points to another. In this process, the pixel burning will start from a seed point of a Region of Interest up until the entire region of interest is covered. After pixel burning, all the pixel points that lie on the edge most part of the burned region will be stored in its knowledgebase as outermost points. Then, bounding box will be generated, tracing the mean of values from the outermost points, making the system capable of drawing tilted bounding box. After generating the bounding box, the direction of tilt and its angle will be determined. Counter rotation of the image will be implemented next, considering the direction detected and the angle computed. Lastly, OCR engine will be used to recognize text characters.

A. Identifying the Direction and Angle of Tilt

First step of the whole process of the system is the text image acquisition. This process is done by a camera. Then, the source image will be subjected to pre processing that includes binarization and edge detection to make the image be in pure black and white and to remove noise. After pre-processing, grass fire algorithm will be implemented. In this process, the algorithm starts the pixel burning at a seed point and then, it will spread out to the entire region of interest that covers that seed point that is why it is called region growing because all the pixel that cover the seed point will be selected as part of a new region [4][5]. Figure 2 illustrates pixel burning.

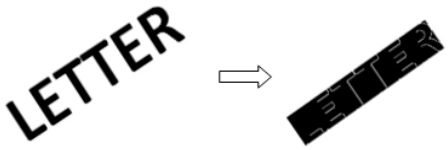


Figure 2 Pixel Burning

The information about the pixel burned are placed on a list or stored in a memory thus making the information about the outermost points be isolated. The system will get the mean of values from the outermost of points firstly generated by the algorithm. Those mean of values will be used to create the straight lines which will lead to generation of the tilting bounding box, which makes it the most essential part of the system. The smallest possible bounding box will enclose the word with all the mean of values of outermost points considered as seen on figure 3. With these tools, the system will be able to create bounding box for the words that are inclined.

Figure 3 Bounding Box

After that, significant points will be derived to be used on detecting and computing the direction and angle of tilt respectively. Since the image is composed of pixels supposedly lying on the Cartesian plane, and the bounding box has been already generated, some information about the bounding box can be established. The bounding box generated is a rectangle consists of two longer sides, two shorter sides and four corner points with x and y coordinates. Significant points will always be the endpoints of the longer side with lower y coordinate as seen on Figure 4.

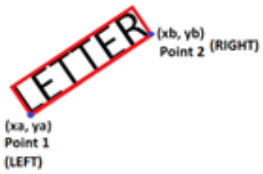


Figure 4 Significant Points

There is a special tilt case that the system will encounter wherein both of the longer side of bounding box has the same lowest Y coordinate. In this case, no significant points will be established; therefore, no direction detection and angle computation will happen because the system assumes that the image is tilted at 90 degrees making it to be subjected immediately to rotation (90 degrees, clockwise). After establishing the significant points, a decision making process seen on figure 5 will be used on detecting the direction of tilt by comparing its y coordinates. Then, a reference triangle will be drawn to get the angle of tilt through the use of the formula 1 shown on figure 6.

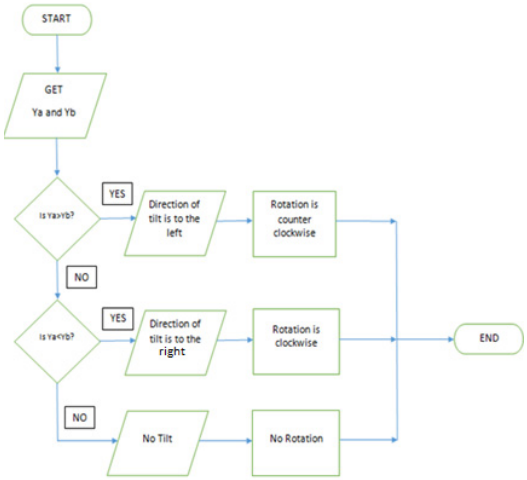


Figure 5 Direction Detection Decision flow

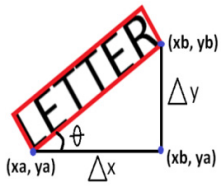


Figure 6 Computation of Angle of Tilt

$$\theta = \tan^{-1}\frac{\Delta y}{\Delta x} \quad (1)$$

After detecting and computing the direction and angle of tilt, counter rotation will be implemented to make the image be back to zero degree orientation. Then, Tesseract Optical Character Recognition will be used together with its confidence function [6]. From the scale of 0-9 with 0 being the best and 9 being the worst, Tesseract OCR engine make judgment on how confident it is that the character recognized is really t he correct character. Then, those values will be fed to formula 2 for the word confidence computation. The system implemented OCR twice thus computing the word confidence also twice, on the word's zero degree orientation and on its 180 degree counterpart as seen on figure 7.

$$Word\ Confidence = \frac{((10-CC1)+(10+CC2)+(10+CC3)+\cdots+(10+CCn))}{10n} \quad (2)$$

CC – Character Confidence

n – Number of Characters in the Word

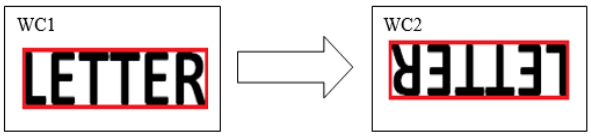


Figure 7 Stages of Word Confidence Reading

After computing the word confidences, it will be used to decide for the output recognition. The output recognition will always be the word with higher confidence.

B. Acquiring the system's reliability on recognition of every rotated text characters in different rotational variations

To determine the reliability of the system, each sample will be subjected in eight different tilt cases to see if there is a significant variation in recognition for each tilt case. Tilt cases are as follows: Case 1 on zero degrees, case 2 on 45 degrees, case 3 on 90 degrees, case 4 on 135 degrees, case 5 on 180 degrees, case 6 on 225 degrees, case 7 on 270 degrees, case 8 on 315 degrees. Correct recognition will tell that the rotation done is right and will be marked as success rotation and recognition. Success rate for each tilt cases will be computed as seen in formula 3.

$$\text{Success Rate(per tilt case)} = \frac{total\ number\ of\ successful\ rotations}{total\ number\ of\ samples} x100 \quad (3)$$

For the overall reliability, the researchers will get the average of all success rates as seen in formula 4.

$$Reliability = \frac{\sum Success\ Rate\ (per\ tilt\ case)}{8} \quad (4)$$

3. Results and Discussion

A. Result of Identifying the Direction and Angle of Tilt

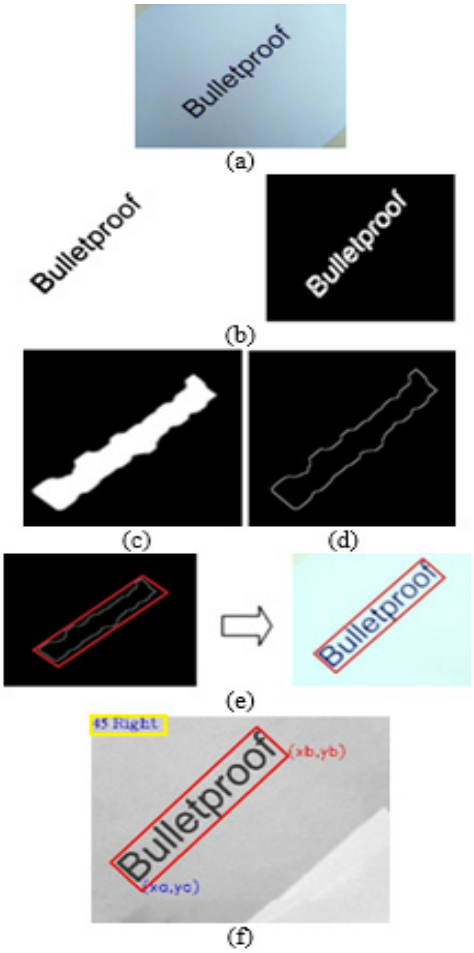


Figure 8 Data Outputs involved in Identifying the Direction and Angle of Tilt (a) Source Text Image (b) Pre-Processed Image (c) Output After Pixel Burning (d) Output After Isolating Outermost Points (e) Bounding Box Generated (f) Direction Detected and Angle Computed

Results are gathered from the 100 samples prepared by the proponents. Figure 8 shows the data outputs of the processes involved in identifying the direction and angle of tilt. Figure 8 (a) shows the source image from one of the samples and (b) shows the output images after pre-processing. Figure 8 (c) shows the output image after pixel burning has been done wherein the white part represents the region of interest burned entirely while Figure 8 (d) shows the output image after the outermost points has been isolated from the region of interest represented by the irregular white line. Figure 8 (e) shows the output images after the bounding box has been generated from the outermost points represented by the red box and Figure 8 (f) shows the output image after the direction and angle of tilt has been identified which, from that specific sample, is 45 degrees to the right written in blue font color. After identifying the direction and angle of tilt, rotation is implemented wherein the image will be rotated as per the angle computed in contrast to the direction identified to make the image be in 0 degree orientation before recognition.

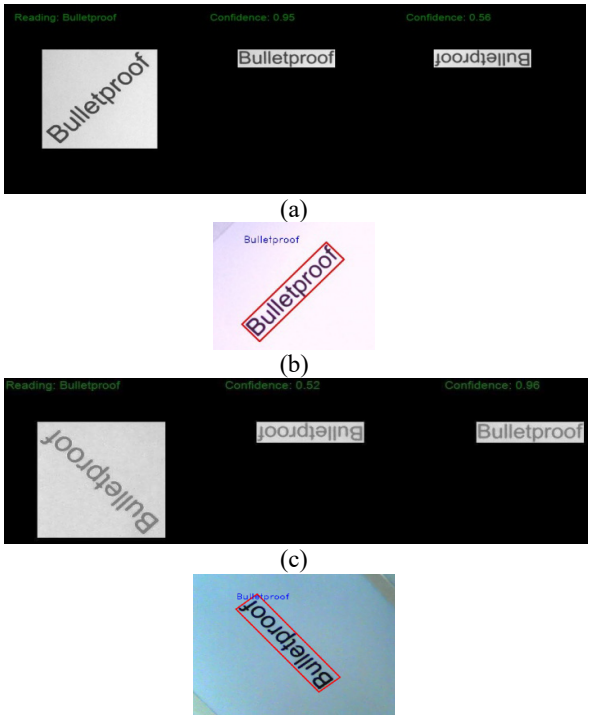


Figure 9 Data outputs After Recognition and Word Confidence Reading (a) WC1 greater that WC2 (b) Output Recognition (c) WC1 less than WC2 (d) Output Recognition

Figure 9 shows the data outputs specific to word confidence reading process as well as the data outputs for recognition after comparing the word confidence readings. Figure 9 (a) shows the stages of word confidence reading when WC1 is greater than WC2. As seen, the reading on first stage is 0.95 while the reading on the second stage is 0.56 making the system decide the output to be on the first stage reading as seen on Figure 9 (b). As seen on figure 9 (c) the reading of word confidence on the second stage is 0.96 which is higher than the word confidence reading on the first which is 0.52, thus, making the output recognition took place on the second stage as seen on figure 9 (d). Figure 10 shows the data outputs of recognition one of the samples subjected to eight (8) different tilt cases.

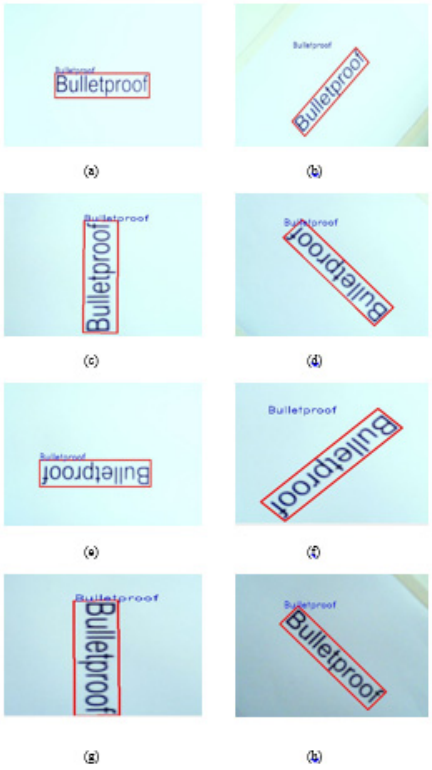


Figure 10 Data Outputs of Recognition for every Tilt Case (a) Case 1; 0 degree (b) Case 2; 45 degrees (c) Case 3; 90 degrees (d) Case 4; 135 degrees (e) Case 5; 180 degrees (f) Case 6; 225 degrees (g) Case 7; 270 degrees (h) Case 8; 315 degrees

B. Acquired system's reliability on recognition of every rotated text characters in different rotational variations

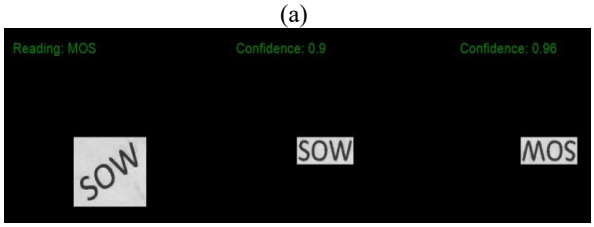
TILT CASES	SUCCESS RATES (%)
1 (0 Degree)	98
2 (45 Degrees)	97
3 (90 Degrees)	95
4 (135 Degrees)	90
5 (180 Degrees)	91
6 (225 Degrees)	90
7 (270 Degrees)	94
8 (315 Degrees)	95
RELIABILITY:	**93.75%**

Table 1 Results of the Study

From 100 samples, success rate is computed in each tilt cases and provided the following outputs: 98% for the first tilt case, 97% for second tilt case, 95% for third tilt case, 90% for the fourth tilt case, 91% for the fifth tilt case, 90% for the sixth tilt case, 94% for the seventh tilt

case, 95% for eighth tilt case. The overall reliability of the system in terms of recognizing every rotated text characters is 93.75%. Table 1 shows the summary of the success rates computed per tilt case and the reliability of the system. As seen, cases 4, 5 and 6 has the lowest success rate due to its first rotated image resulting always to 180 degrees orientation. When optical character recognition is implemented, it calculates the word confidence on the first rotation making the higher chance of getting a higher word confidence than the second word confidence reading. An error rate of 6.25% was also determined. The error rate is consists of word misjudgment and word rearrangement errors. Figure 11 shows the data outputs for misjudgment error. As seen on figure 11 (a) the sample "SOW" was recognized as "MOS". This means that the recognition took place on the 180 degree orientation of the word because the 180 degree orientation of the word formed another word with higher word confidence value than the original word that confuses the system. Figure 11 (b) shows the word confidence on two stages of rotation. As seen on the figure, the sample on the first rotation has an output word confidence of 0.9 and an output on second rotation of 0.96 which is higher than the first reading. Because of that, the output of the system is the second stage of recognition which is the 180 degree counterpart of the sample. This kind of error is usually present on some word, and on some tilt cases depending on the combination of the characters inside the word. The error sometimes happened due to the varying rotation, and sometimes, due to the combination of the characters in the word solely. This error frequently happened on cases 4, 5, and 6.

(a)



(b)

Figure 11 Data Outputs for Misjudgement Error (a) Word Confidence Reading (b) Output Recognition

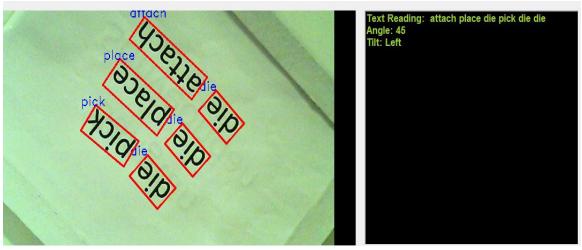


Figure 12 Data Output for Word Rearranging Error

Figure 12 shows the data output for word rearrangement error. As seen, the sample "die pick die place die attach" were rearranged during recognition when subjected to rotational variations becoming "attach place die pick die die". This happens because the system is reading the recognized word from top to bottom, left to right, disregarding the arrangement of the words when subjected to rotation.

4. Conclusion

The researchers developed a system that can recognize text in different rotational variations by obtaining the right orientation of the text through acquisition of the direction and angle of tilt using geometric and trigonometric principles. Even though the reliability of the system is high, there are still some incidents that the system fails to recognize the text properly. First is when a word, when rotated 180 degree, will result to combination of new characters forming another word, causing the confusion of the system in choosing between the two words from the recognition of two rotated images. This happens frequently on cases 4, 5, and 6 because of the fact that the word being process first is upside down. Second is caused by multiple line of words that when subjected to rotational variations, words were being rearranged, producing an output of disordered words. This happens because the system read the recognizes words from top to bottom. Regardless of the mentioned incidents where errors occurred, the proponents conclude that the system can still detect and recognize text in different rotational variations.

5. Acknowledgements

The proponents wish to express their deepest appreciation and respect to the following who gave possibility to complete this study, without them this could not have been done.

Firstly, the proponents would like to express their deepest faith and gratefulness to the Almighty God whose presence is always been there to guide them and continuously bless them.

Second, to their beloved families and relatives, who support and encourage them unconditionally in spite of their flaws, and are still there to provide for them financial, moral support, lots of love and boundless understanding.

Third, to their Thesis adviser, Engr. Roselito E. Tolentino, for his absolute support, advices, guidance, comments, suggestions, and provisions that benefited them in the completion and success of this study.

Lastly, the proponents would also like to extend their gratitude to their friends and families who have given and shared with them their laughter during times of frustration

References

[1] Kaur, T. and Neeru, N. Text Detection and Recognition from Natural Scene. 2015

[2] Moeslund, T. B. Introduction to Video and Image Processing: Building Real Systems and Applications. 2012

[3] Neumann and Matas. Real-time Lexicon-free Scene Text Localization and Recognition. 2015

[4] Shilpa K.M., Shankar R., Suma Swamy. Text-Line Detection, Segmentation and Recognition In Natural Scene. 2017

[5] Smith, Ray. An Overview of the Tesseract OCR Rngine. 2007

[6] Yun-Kyoo Ryoo, Chan-Myeong Han, Ja-Hyo Ku, Dae-Seong Jeoune,and Young-Woo Yoon. Grassfire Spot Matching Algorithm in 2-DE. International Journal of Bio-Science and Bio-Technology. 2013

Biographies

Marron Jann M. Anyayahan was born in Manila, Philippines on January 5, 1996. He is currently taking up Bachelor of Science in Electronics Engineering at Polytechnic University of the Philippines – Santa Rosa Campus. He was a part of Excelitas Technologies Philippines, Inc. at Cabuyao, Laguna in 2015 and Asian Vision Cable Holdings Inc. at Makati City in 2016 for his On the Job Training. Marron Jann M. Anyayahan is currently a member of Association of Electronics and Communications Engineering Students and was a member of Institute of Electronics Engineers of the Philippines Laguna Chapter.

Ma. Kelaine B. Balinas was born in Carmona Cavite, Philippines on June 13, 1996. She is currently taking up Bachelor of Science in Electronics Engineering at Polytechnic University of the Philippines, Santa Rosa Campus. She was a part of Amkor Technology Philippines at Binan, Laguna in 2015 and Total Power Box Solutions, Inc. at Silang, Cavite in 2016 for her On the Job Training. Ma. Kelaine B. Balinas is currently a member of Association of Electronics and Communications Engineering Students and was a member of Institute of Electronics Engineers of the Philippines Laguna Chapter.

Aldrin P. La Madrid was born in Sta. Rosa Laguna, Philippines on February 18, 1996. He is currently taking up Bachelor of Science in Electronics Engineering at Polytechnic University of the Philippines, Santa Rosa Campus. He was a part of GF Micro Optics Philippines, Inc. at Binan, Laguna in 2015 and PLDT Las Pinas – Alabang Customer Service Operations Zone at Muntinlupa, Metro Manila in 2016 for his On the Job Training. Aldrin P. La Madrid is currently a member of Association of Electronics and Communications Engineering Students and was a member of Institute of Electronics Engineers of the Philippines Laguna Chapter.

Mark Anthony T. Laurel was born in Manila, Philippines on September 15, 1995. He is currently taking up Bachelor of Science in Electronics Engineering at Polytechnic University of the Philippines, Santa Rosa Campus. He was a part of Cirtek Electronics Inc. at Binan, Laguna in 2015 and PLDT Calamba Branch at Calamba, Laguna in 2016 for his On the Job Training. Mark Anthony T. Laurel is currently a member of Association of Electronics and Communications Engineering Students and was a member of Institute of Electronics Engineers of the Philippines Laguna Chapter.

Christelle Joy D. Lopez was born in Sta. Rosa, Laguna, Philippines on June 6, 1996. She is currently taking up Bachelor of Science in Electronics Engineering at Polytechnic University of the Philippines, Santa Rosa Campus. She was a part of Sensor Scientific Phils., Inc. at Calamba, Laguna in 2015 and Total Power Box Solutions, Inc. at Silang, Cavite in 2016 for her On the Job Training. Christelle Joy D. Lopez is currently a member of Association of Electronics and Communications Engineering Students and was a member of Institute of Electronics Engineers of the Philippines Laguna Chapter.

Roselito E. Tolentino is a registered Electronics Engineer and IECEP-Member. He is a graduate of B.S. Electronics and Communication Engineering at Adamson University in 2004 under the scholarship of DOST-SEI. He finished his Master of Science in Electronics Engineering Major in Control System at Mapua Institute of Technology under the scholarship of DOST-ERDT. He currently takes up Doctor of Philosophy in Electronics Engineering at the same Institute. He is currently working as a part time instructor at Polytechnic University of the Philippines Santa Rosa Campus and De La Salle University - Dasmarinas. His research interests are more on Robotics and Machine Vision.